4/pro

Method for calculating a transmission window size

Technical field of the invention

The invention relates to a method for selecting a window size for a packet switched connection between a first and a second party. The associated window is used by a sending party for a window based congestion control mechanism for avoiding or handling congestion on a communication path. The window size defines the maximum number of data packets that can be sent by a sending party before an acknowledgement of the reception of a packet is received by said sending party.

Description of related art

Communication systems using window based congestion control are well known as for example systems operating according to a TCP/IP (Transmission Control Protocol / Internet Protocol) or systems operating according to a SCTP (Stream Control Transmission Protocol). Such systems permit the sending of a certain number of packets from a sender to a receiver, before an acknowledgement of a reception of a packet is received at the sender. The number of packets that may be sent unacknowledged is called window size. As multiple packets are sent before the reception of an acknowledgement for at least one of the packets, the efficacy of the use of a transmission channel is improved. In general, a larger window size increases the utilisation of transmission resources.

However, a fixed larger initial window bears the risk that the window is too large in some situations, where congestion occurs somewhere in the network. In such situations a too large window contributes on top to the congestion, which could lead potentially to problems.

Besides the initial window, TCP uses some more window sizes to reinitialise the congestion window after certain events. The definitions of the window sizes are described in M. Allman, V. Paxson, W. Stevens: TCP Congestion Control, RFC2581, published April 1999 as IW (initial window), which is the size of the sender's congestion window after the three-way handshake is completed, LW (loss window), which is the size of the congestion window after a TCP sender detects loss using its retransmission timer, and RW (restart

window) as the size of the congestion window after a TCP restarts transmission after an idle period. According to M. Allman, V. Paxson, W. Stevens, the initial window can be either 1 or 2 segments. They define a loss window size of 1 segment and a restart window size that should have the same value as the initial window size.

M. Allman, S. Floyd, C. Partridge suggest a different initial window size in:" Increasing TCP's Initial Window", RFC2414, published September 1998. According to them, the initial window size can be set to

Initial window= min (4*MSS, max (2*MSS, 4380 bytes))

wherein MSS is the maximum segment size, min is the minimum function and max the maximum function.

Applying this to the most popular MSS (Maximum Segment Size) yields an initial window of 4 segments for 512 or 536 bytes while 3 segments for 1460 bytes.

Increasing a TCP sender's IW, LW, and RW can significantly improve end-to-end performance but also bears the risk of causing network congestion. This is why it is strongly discouraged to do this when running TCP over the wide-area Internet. However, when running TCP only across a "private" network, e.g., when running it between a proxy and a mobile terminal across a wireless access network, one could afford large values of IW, LW, and RW. Nevertheless, if congestion can also occur within that "private" network, it is problematic to find optimal values for IW, LW, and RW.

H. Balakrishnan, S. Seshan introduce in "The Congestion Manager", RFC3124, published June 2001 the usage of known window sizes for the set up of a new connection. However, this mechanism is limited to the use in the case that the sender and receiver of the new connection are the same as of an ongoing connection.

Currently there is no mechanism that enables to adapt the window size for a packet switched connection in a way that avoids the waste of transmission resources.

Therefore it is object of the invention to provide a method and means for implementing the method that enables to adapt the window size for a packet switched connection in a way that avoids the waste of transmission resources.

This is solved by the method of claim 1, the window size selecting unit of claim 19 and the threshold value determining unit of claim 22.

It is advantageous that window sizes are determined by using information about the pipe capacity of a connection the window will be used for. By this a more appropriate window size can be determined thus increasing the utilisation of transmission resources. It is further advantageous that an upper threshold value is determined for window sizes. An increase of window sizes above the upper threshold value would lead to packet losses and by that to less efficient use of transmission capacities.

Further advantageous embodiments can be derived from the dependent claims.

Summary

The invention introduces a solution that is applicable to any end-to-end protocol that uses window-based congestion control. In particular, it applies to TCP, but also to SCTP (Stream Control Transmission Protocol).

The invented method makes IW, LW, and RW adaptive to the communication network. This is especially valuable for, but not limited to, communication networks comprising an air interface. The maximum bit rate on an air interface varies strongly while transmission capacity on the air interface is expensive. The invented method can be used to calculate IW, RW and LW together if they are set to equal values or each of them may be selected separately. The invented method is used to select window sizes based on the pipe capacity of a connection, the destination of a connection and the loss history of a connection. Furthermore the loss history of connections with the same pipe capacity or with a pipe capacity that falls into the same predefined range of pipe capacities can be taken into account.

Therefore a method for selecting a window size for a communication system for connecting a first and a second party is introduced. The communication system comprises means for setting up a packet switched connection between the parties, wherein a sending party is adapted to use a window based congestion control mechanism for avoiding or handling congestion. The window defines the maximum number of data packets that may be sent by a sender before an acknowledgement of the reception of a packet is received by the sender. The following steps are performed when executing the method:
retrieving information about a bit rate of a link belonging to a path across which the connection between the parties is set up, retrieving information about an estimation of a round trip time on the connection between the parties, determining an estimation of a pipe capacity for the connection between the parties according to the retrieved bit rate and the estimation of the round trip time of the connection, determining an upper threshold value for the window size based on the pipe capacity, and selecting a window size value above zero and below or equal to the upper threshold value.

The invented method can comprise the additional steps of storing the selected window size together with an indication of the pipe capacity, or a predefined range of pipe capacities comprising the pipe capacity, of the connection. The storing of the selected window size has the advantage that a selected window size can be used for further connections. The storing of the pipe capacity or a predefined range of pipe capacities comprising the pipe capacity, of the connection has the advantage that a stored window size can be selected depending on the pipe capacity.

The invented method can further comprise the step of determining a destination of the connection. In that case, the selected window size is stored together with an identification of said destination. This enables to select a stored window size depending on the pipe capacity and the destination of a connection. If the communication system is a cellular communication system, a destination is one of a location area, a routing area, a cell, a service area or an area served by a radio network controller, a mobile services switching centre, a radio base station, or a serving general packet radio service support node.

In an embodiment of the invention, the communication system is a cellular communication system and the link is a wireless link.

In a further embodiment of the invention, the window is one of an initial window, a loss window or a restart window.

In the method and its embodiments, a party may be one of a proxy server, a mobile user equipment, a radio network controller, a general packet radio service support node, a radio base station, and a fixed network terminal.

In a preferred embodiment of the invention, the upper threshold value is in a range of plus or minus two packets around twice the pipe capacity or twice the higher value of the predefined range of pipe capacities comprising the pipe capacity of the connection the window is used for.

The method and its embodiments may comprise the additional steps of receiving a congestion indication for a connection before an acknowledgement for all packets sent in an initial window, a loss window, or a restart window is received, and of selecting a smaller window size. In a preferred embodiment of the invention, the selected smaller window size is about half the size of the window size used before, unless the former window size was one.

The method and its embodiments may comprise the additional step of detecting an increase of the pipe capacity of a connection, and selecting a new window size for said connection, wherein the new window size is one of an initial window size, a loss window size or a restart window that are used for connections with the same pipe capacity or with a pipe capacity that falls into the same predefined range of pipe capacities as the increased pipe capacity, or wherein, if none of said window sizes is available, a value is selected for the new window size that is n times the increased pipe capacity, with n greater than or equal to 1 and smaller than or equal to 2. An appropriate upper threshold value for the new pipe capacity that allows increasing a congestion window up to the selected window is determined and used. In an embodiment of the invention, a congestion window used for the

connection is set to the selected window size. In a preferred embodiment a slow start threshold value for the connection is set to said selected window size.

The invented method and its embodiments may also comprise the additional steps of monitoring for a predefined number of seconds or number of connection set-ups or restarts that no congestion indication is received for a connection before an acknowledgement for all packets sent in an initial window a loss window or a restart window is received, and selecting a larger window size that is smaller than or equals the upper threshold value. In a preferred embodiment, the selected larger window size differs from the window size used before by a predefined constant number.

In an embodiment of the invented method, the monitoring and the selecting of a larger window size are performed separately for different destinations.

In a preferred embodiment of the invention, the selected window size is used for a further connection with the same destination and the same pipe capacity or with a pipe capacity that falls into the same predefined range of pipe capacities that is set-up, restarted or wherein a packet loss was detected. That is connections with the same pipe capacity or with a pipe capacity that falls into the same predefined range of pipe capacities are treated as a group and that connections that belong to said group have the same IW, LW and RW.

The invention further relates to a window size selecting unit for a communications system for connecting a first and a second party, wherein a sending party is adapted to use a window based congestion control mechanism for avoiding or handling congestion on a communication path. The window is defining the maximum number of data packets that may be sent by a sender before the sender receives an acknowledgement of the reception of a packet. The window size selecting unit comprises an input/output unit for sending and receiving data, a processing unit for controlling the other units, and is characterised by a selection unit for selecting a window size above zero and below or equal to an upper threshold value for a connection between the parties.

In an embodiment of the invention, the window size selecting unit further comprises a

storage for storing window sizes together with an information about a pipe capacity and a comparing unit for comparing stored pipe capacities and determined pipe capacities.

The window size selecting unit may further comprise a destination determining unit for determining a destination of a connection, wherein the storage is adapted to store an identification of a destination together with the window size and the information about a pipe capacity, and wherein the comparing unit is adapted to compare stored destinations and determined destinations.

The invention also relates to a threshold value determining unit that comprises an input/output unit, a pipe capacity determining unit for determining an estimation of a round trip time of a connection and a bit rate of said connection, and for determining the estimation of the pipe capacity of said connection from the estimation of the round trip time and the bit rate, and a processing unit for controlling the units and calculating an upper threshold value for further use in a window size selecting unit.

Brief description of the figures

Figure 1 depicts a schematic of a communication path between a first and a second party.
Figure 2 depicts a flow chart describing the invented method.
Figure 3a depicts a flow chart describing a section the invented method.
Figure 3b depicts a flow chart describing a further section the invented method.
Figure 3c depicts a flow chart describing a further section the invented method.
Figure 3d depicts a flow chart describing a section of a preferred embodiment of the invented method.
Figure 3e depicts a flow chart describing a further section of a preferred embodiment of the invented method.
Figure 3f depicts a flow chart describing a preferred embodiment of the invented method.
Figure 3g depicts a flow chart describing additional steps for an embodiment of the invented method.
Figure 4 depicts a window size selecting unit.
Figure 5 depicts a threshold value determining unit.

Detailed description of the invention

In the following the invention will be described by means of figures and embodiments. The invention will be explained by using a network comprising a mobile network without restricting the invention to such implementation.

Figure 1 depicts a schematic of a communication path between a first party UE1 and a server S1. The server is connected via a link L11 to an IP based network IP1. Said IP based network is connected via a link L12 to a proxy server P1. Said proxy server is used to connect the fixed connected domain comprising the before mentioned components with a wireless domain via a link L13. The wireless domain comprises the network for mobile telecommunications RN1 and the first party UE1. The network for mobile telecommunications RN1 is connected to the proxy via said link L13. It is further connected to the first party UE1 via a radio link RL1. The proxy P1 acts as a party towards the server S1 and the first party UE1. In the following the connection radio link RL1, radio network RN1 and link L14, between the first party UE1 and the proxy P1 acting as a second party is regarded. The invented method is used to determine a window size for said connection. Figure 2 depicts a flow chart describing the invented method. After starting 201 the method a first optional step 202 is performed. At that step the proxy P1, acting as a window size determining unit, categorises all mobile terminals UE1 that currently terminate at least one active TCP (Transmission Control Protocol) flow at said proxy P1 into destinations, according to the location of the mobile terminal. Instead of TCP the invented method can be executed for any window based packet transmission protocol as for example SCTP or DCCP (Datagram Congestion Control Protocol).

In a step 203, the proxy, again acting as a window size determining unit, groups all TCP flows, with the same pipe capacity into the same group. In a preferred embodiment, TCP flows with a pipe capacity that falls into the same predefined range of pipe capacities are grouped into the same group. Said range is defined for example by operator settings or by a vendor of a computer program that controls the window size determining unit in a way that it executes the invented method. Step 203 is run separately and independently for those

active TCP flows that terminate at the same destination. If the optional step 202 has not been performed, step 203 is run separately and independently for all active TCP flows.

At step 204, for all flows with the same pipe capacity or with a pipe capacity that falls into the same predefined range of pipe capacities a window size is determined. Step 204 can be performed several times until for each TCP flow a window size is determined. In an embodiment of the invention a window is one of an initial window, a loss window or a restart window.

Step 202 is described in more detail by means of figure 3a. After starting step 202 in the sub-step startddest, the destination of a connection is determined in the sub-step ddest. This can be performed for example by gaining information from the radio network. Depending on which information from the mobile network is available to the window size determining unit and a preferred granularity, a destination can for example be one of a location area, a routing area, a cell, a service area or an area served by a radio network controller, a mobile services switching centre, a radio base station or a serving general packet radio service support node. After determining a destination for each connection the method shall be performed for, the step 202 is ended in the sub-step endddest. Alternatively the sequence of steps or each of the steps 202, 203 and 204 can be performed for a single destination, a group of destinations or all destinations. An advantage of this step 202 is that mobile terminals of the same destination share the same potential bottleneck link in the mobile network, and that different destinations have a different potential bottleneck link. Thus, it can be expected that mobile terminals of the same destination with the same potential bottleneck link share some transmission characteristics.

Step 203 is depicted in more detail in figure 3b.When step 203 is started in sub-step startdpcap, the estimation of the round trip time of the connection RL1, RN1, L13 between the parties P1, UE1 is determined in a sub-step drtt. The Round-trip-time is estimated for example based on knowledge about the network or experience collected on said network or compatible networks. In a further sub-step dbrate, the bit rate is determined of a link L13, RL1 belonging to a path across which the connection between the parties is set up. The pipe

capacity of a link is the minimum number of bytes a sending party needs to have in flight to fully utilize its available bandwidth. It can be calculated as the product of bit rate and round trip time in the sub-step dtcap. Afterwards the step 203 ends in the sub-step enddpcap. In a preferred embodiment of the invention, the bit rate on the bottleneck link is determined for the estimation of the pipe capacity. In the depicted connection this is the radio link RL1. Thus, the pipe capacity is simply the product of the radio bearer RL1 bit rate and the round-trip delay between the proxy P1 and the mobile terminal UE1. It is known to a person skilled in the art that the proxy P1 can attain knowledge about the mentioned bit rate and round-trip delay associated with a specific TCP connection. For example, on request from the proxy P1 the network for mobile telecommunications RN1 could signal that information to the proxy P1, or the proxy P1 could have access to a profile database where that information is kept.

In an embodiment of the invention, connections with the same pipe capacity are grouped.

In a preferred embodiment of the invention, in order to reduce the number of executions of the invented method, not only connections with exactly the same pipe capacity are treated equally, but also connections within a predefined range of pipe capacities.

The figures 3c, 3d, and 3e are used to describe step 204 in more detail. The embodiment of step 204 as depicted in figure 3c comprises the sub-steps of starting the step startselwin, of determining an upper threshold value for a window size dupthresh, of selecting a window size, and of ending the step endselwin. The upper threshold value of a window size is determined as twice the pipe capacity of the connection the window is used for. A window size above twice the pipe capacity does not increase the performance of a connection. In the next sub-step selwin a window size is determined. Said window size has a value above zero and below or equal to the upper threshold value. In a preferred embodiment of the invention, the value is higher than the pipe capacity of the connection. The higher the value, the smaller the loss of transmission capacity, but the risk of congestion or of losing packets increases.

Figure 3d depicts an embodiment of step 204, with the additional sub-step store selected window size sselwin. The selected window size is stored to be reused for the same connection if a packet is lost. In an embodiment of the invention, the stored window size is stored together with an indication of the pipe capacity or the range of pipe capacities the connection belongs to and is used for another connection with the same pipe capacity or within the same predefined range of pipe capacities. In a preferred embodiment of the invention, the stored window size is stored together with an indication of the pipe capacity or the range of pipe capacities the connection belongs to and an identification of the destination for the connection, and the window size is used for another connection with the same pipe capacity or with a pipe capacity that falls into the same predefined range of pipe capacities only if it has the same destination.

Figure 3e depicts an embodiment of step 204, with the additional sub-steps of receiving an indication of a packet loss recvpktloss and of selecting a new, smaller window size selswin. In the sub-step of receiving an indication of a packet loss recvpktloss, an indication is received that a packet of an initial flight was lost. An initial flight is a number of packets send in a first window after a set-up or a restart of a connection. If one of the packets sent in an initial flight is lost, congestion can be assumed. Therefore, a new, smaller window size is selected in the sub-step selswin. In a preferred embodiment of the invention, the new window size is half the former window size unless the window size is already one maximum segment size. In the following the size of a window is measured in multiples of a maximum segment size to make it easier for a person skilled in the art to understand the invention. In a preferred embodiment of the invention the new selected window size is stored and used as described by figure 3d.

Figure 3f depicts an embodiment of the invented method with the additional steps of determining an increase of pipe capacity for a connection dipcap, selecting an increased window size for the connection seliwin, and of introducing the increased window size for the connection intseliwin. In the case that a pipe capacity of a connection is increased, for example because a radio link receives more bandwidth, an indication is sent to a window size selecting unit. The window size selecting unit selects a new window size for a

congestion window in a step seliwin. A congestion window defines the number of packets that may be sent before an acknowledgement is received at the sender. The congestion window is set to the size of a loss window after a packet loss, of an initial window when a connection is set up, or of a restart window when a connection is restarted. During an active for example TCP connection, the congestion window size varies. It should be noted that the change of a window size changes the size of the first congestion window after a set-up of, restart of or packet loss on a connection. In the following embodiments however, the size of a congestion window is changed in the latter use of a connection. In a preferred embodiment the congestion window size is increased linearly until either an upper threshold value is reached or a congestion indication is received. If a congestion indication is received, the congestion window size reduced to about half its former value. At the step seliwin, the window size selecting unit determines whether there is already a window size stored for connections of the same pipe capacity or within the same range of pipe capacity as the increased pipe capacity. If so, the stored window size will be used for the congestion window size. In a preferred embodiment stored value will be used only if the identification of the stored destination for the value matches with an identification of the destination of the connection. If no stored window size is available, the new window size is selected as a value that is a multiple of the new pipe capacity. In a preferred embodiment of the invention, the new window size is in a range between the increased pipe capacity and twice the increased pipe capacity.

The new window size is introduced for the use for the connection in a next step intselwin. In a first sub-step the upper threshold value for a window size is set to twice the increased pipe capacity plus or minus two maximum segment sizes. Three alternative embodiments are introduced for introducing from there on in more detail. In a first and preferred embodiment of introducing the new congestion window size, the slow start threshold value of the connection is set to the selected window size value. This leads to a faster than linear increase of the congestion window size used for the connection. In a second embodiment, the congestion window used for the connection is set to the selected window size. By this, the new congestion window size is used immediately for the connection. In a third

embodiment, no further action is taken which leads to a linear increase of the congestion window size.

Figure 3g depicts a sequence of additional steps that are implemented in an embodiment of the invented method. In a first step startmoni the sequence is started. In a next step moni connections with the same pipe capacity or with a pipe capacity that falls into the same predefined range of pipe capacities are monitored whether a congestion indication is received for an initial flight, a restart flight or a loss flight. A flight is a number of packets send within a congestion window. If the flight is the first flight sent after a set-up of a connection, it is called initial flight and the number of packets is related to the initial window size. If the flight is the first flight sent after a restart of a connection, it is called restart flight and the number of packets is related to the restart window size. If the flight is the first flight sent after a packet loss on a connection, it is called loss flight and the number of packets is related to the loss window size. The monitoring is performed for a certain predefined time interval or for a predefined number of connections set-ups or restarts. If the time for the monitoring expires or the predefined number of connection set-ups or restarts is reached, the monitoring is terminated. It is then assumed that the window sizes can be increased for future set-ups or restarts. Therefore increased new window size values are determined in a next step sellwin. In a preferred embodiment of the invention, the window size is increased by a predefined constant value. The sequence of additional steps in ended in a step endmoni.

Figure 4 depicts a window size selecting unit WSSU4. Said unit comprises an input/output unit IO4 for receiving and sending data, a processing unit PU4 for controlling and coordinating the other units, a selecting unit SU4 for selecting a window size, a store ST4 for storing window sizes, a comparing unit CU4 for comparing stored pipe capacities and determined pipe capacities or the respective predefined ranges, and a destination determining unit DDU4. The units comprised in the window size selecting unit WSSU4 can be implemented as depicted in a single housing or may be distributed within a node or even among several nodes. The units may be realised by means of hardware of software or a combination of both. In an embodiment of the window size selecting unit WSSU4 a

destination determining unit DDU4 is optional. In another embodiment of the invention, the comparing unit CU4 is adapted to compare stored destinations and determined destinations.

In an embodiment of the invention, the initial window, the loss window and the restart window are of the same size.

After booting of an entity that is adapted to act as a sending party, initial values are set for the initial window, the loss window and the restart window in said entity. An embodiment is to choose the pipe capacity as this initial value. A preferred embodiment is to choose the twice the pipe capacity as this initial value.